Ensemble Kalman Filter Data Assimilation in the POP Ocean Model

William Casper and Balu Nadiga wcasper@u.washington.edu, btnadiga@lanl.gov

Abstract

We present results from our initial attempts at ensemble-based data asimilation in a global ocean general circulation model (POP, Parallel Ocean Program) at one degree resolution. We hypothesize that model error (due to unresolved scales) is responsible for the insufficient ensemble diversity and spread witnessed in initial tests. To correct this, we implement a spread restoration algorithm which restores estimates of the forecast error covariance toard a fixed error covariance estimate taken from a control run. Initial test show significant improvement in the assimilation system over both uncorrected runs and runs using adaptive inflation schemes.

Background

Data assimilation (DA) routines are methods of introducing real-world observations into numerical models to more accurately approximate a physical system's "true state".

Statistical DA methods use the numerical model to advance an initial state to a time t_j where observation information exists, producing a **forecast state** $\mathbf{x}_f$. The observations are then assimilated, producing an **analysis state** $\mathbf{x}_a$, and the process is repeated until observations at all times are assimilated.

The analysis state $\mathbf{x}_a$ is determined by an **optimal increment** matrix $\mathbf{W}$ which minimizes the expectation value of the analysis error:

$$\mathbf{x}_a = \mathbf{x}_f + \mathbf{W}(\mathbf{y}_o - \mathbf{H}\mathbf{x}_f) \quad (1)$$

$$\mathbf{W} = \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} \quad (2)$$

Here, $\mathbf{R}$ and $\mathbf{B}$ are the observational and forecast error covariance matrices, $\mathbf{y}_o$ is the vector of observations, and $\mathbf{H}$ transforms from the physical space to the observation space.

Ensemble Kalman Filter

An **ensemble Kalman filter** (EnKF) differs from other statistical DA routines in that it uses information from an ensemble of forecasts to estimate the forecast error covariance $\mathbf{B}$.

For POP, we used NCAR's Data Assimilation Research Testbed (DART), an ensemble average Kalman filter (EAKF), and assimilates observations sequentially, one at a time. For each observation y_o, DART assumes ensemble perturbations approximate a Gaussian distribution about the ensemble mean at the observation $\overline{y_f}$ with standard deviation (spread) σ_f. DART then updates the ensemble state to a new Gaussian distribution ($\overline{y_a}$,σ_a using Baysian statistics with the corresponding state variables $\mathbf{x}_a^{(k)}$ of each ensemble member k via linear regression

$$\frac{1}{\sigma_a^2} = \frac{1}{\sigma_o^2} + \frac{1}{\sigma_f^2} \qquad \frac{\overline{y}_a}{\sigma_a^2} = \frac{y_o}{\sigma_o^2} + \frac{\overline{y}_b}{\sigma_b^2} \quad (3)$$

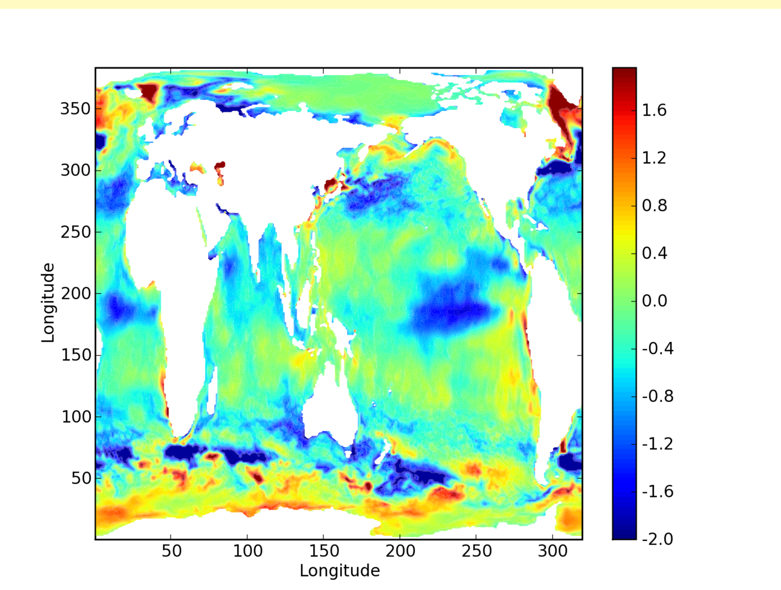
$$\mathbf{x}_a^{(k)} = \overline{\mathbf{x}_a} + \beta(\mathbf{x},y)(\overline{y}_a - \overline{y}_b) + \beta(\mathbf{x},y)\left(\frac{\sigma_a}{\sigma_b} - 1\right)\left(\mathbf{H}\mathbf{x}_a^{(k)} - y_o\right) \quad (4)$$

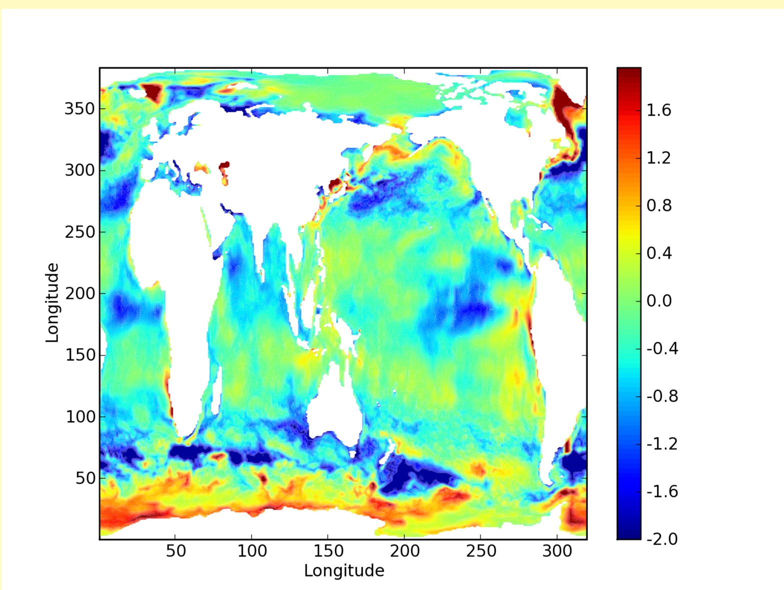
where $\beta(x,y)$ is the linear regression factor between state variable x and observation y.

Experimental Setup

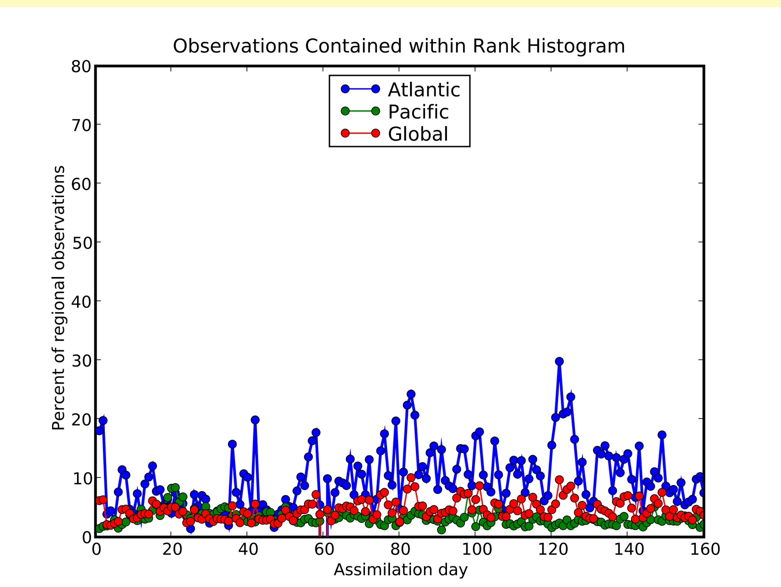
Assimilation uses LANL POP with DART, forced with corrected COREv2 interannual forcing and obs from the World Ocean Database (WOD), starting Jan 1, 1990. Ensemble initialized with states from Jan 1 of different years of a normal-year spinup run. Compared against run without DA.

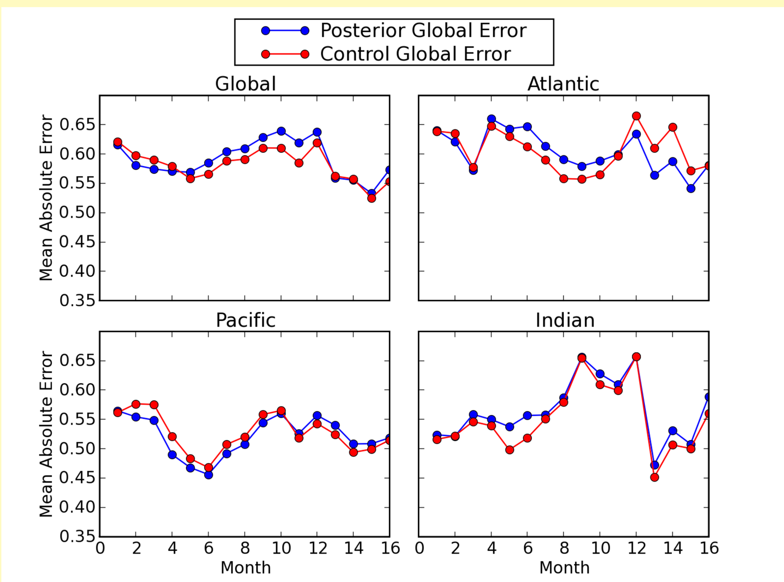
Initial Results





LEFT: control run SST bias **RIGHT:** assimilation run with DART adaptive inflation. Exhibits cold bias in tropics and midlatitudes and warm bias at high latitudes and upwelling regions. No significant improvement with DA.





LEFT: percentage of observations contained within rank histogram ($< 10\%$) **RIGHT:** area-weighted mean absolute error of the monthly-averaged SST anomaly for Jan 1990 through Apr 1991. Exhibits no net reduction in error.

Inflation

Problem: The initial results of data assimilation show very poor assimilation performance.
Hypothesis: Poor performance is due to

- insufficient ensemble spread due to the small ensemble size and diversity and insufficient model variability causing EnKF to underestimate forecast error covariance
- significant model bias causing the ensemble mean to be far from the observation values

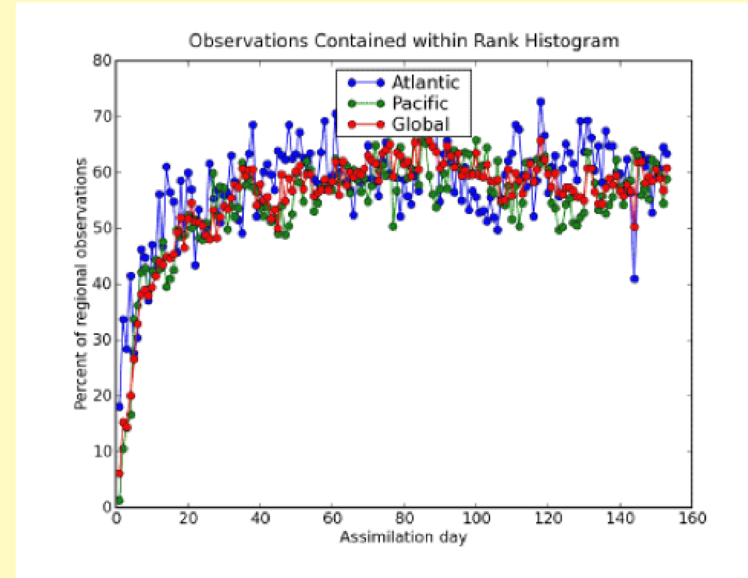
Solution: To help correct for model bias, we employed a simple bias correction scheme (which we do not analyze here). To correct for insufficent spread, we substitute the DART inflation algorithm with a new inflation restoration scheme which restores the forecast ensemble spread S toward a fixed (time-invariant) estimate of the background variability σ obtained from a control run

$$S \mapsto (1-c)S + c\sigma. \quad (5)$$

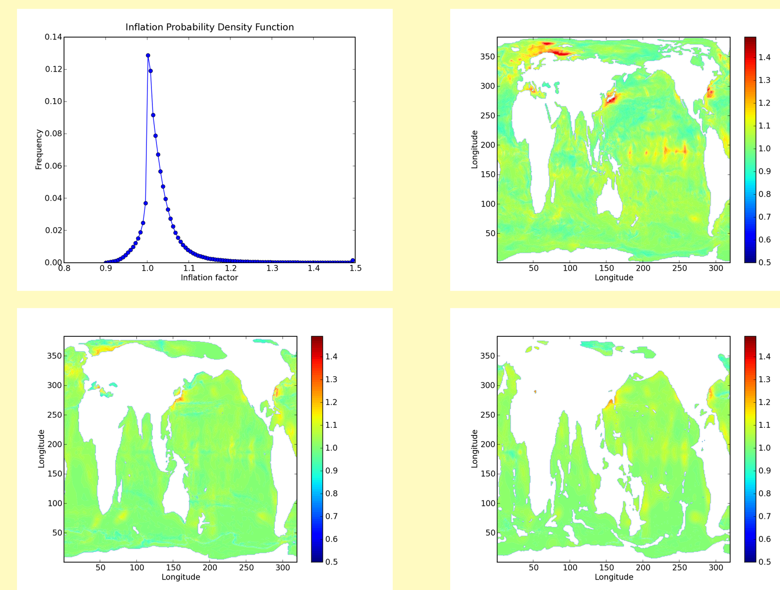
This is analogous to hybrid methods (e.g. EnKF/3D-Var hybrid), which boost the rank of the forecast error covariance matrix.

Improved Results

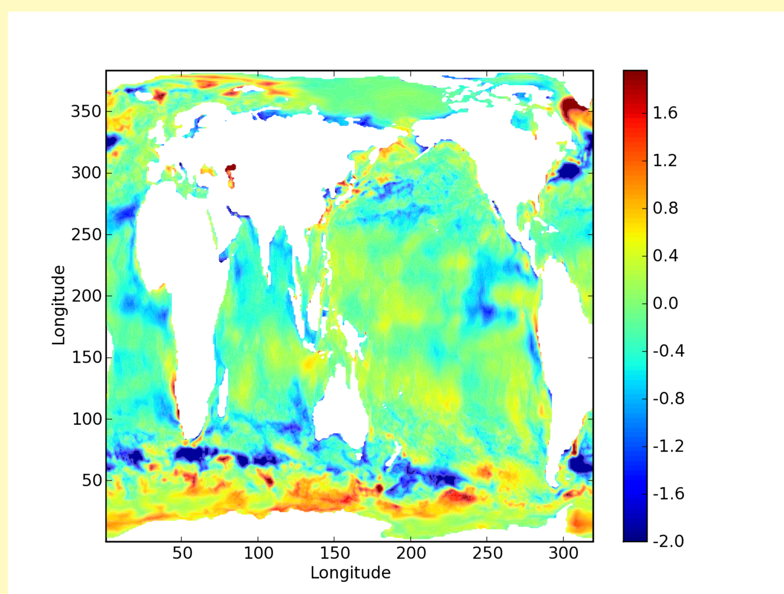
BELOW: The number of observations contained within the rank histogram increases greatly.

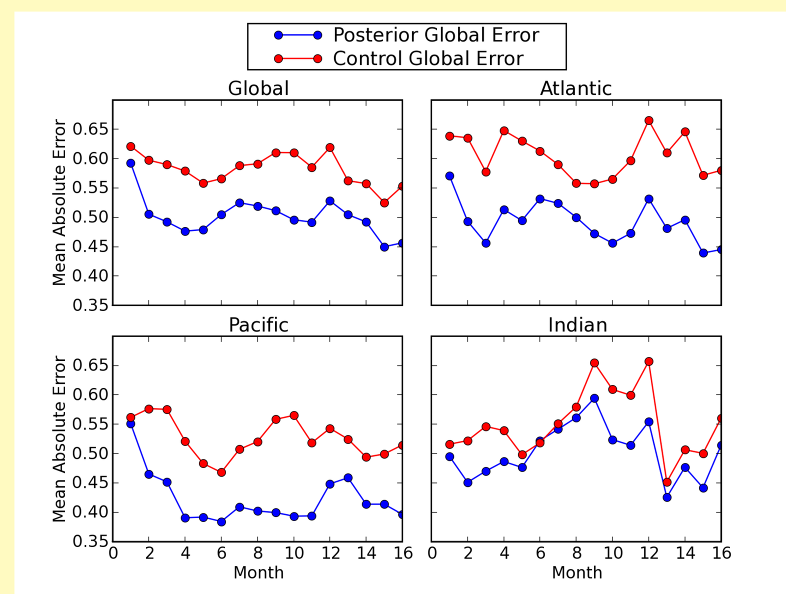


BELOW: Avg. inflation factors for pot. temp. at surface (top right) and depths of 1000 m (bottom left) and 3200 m



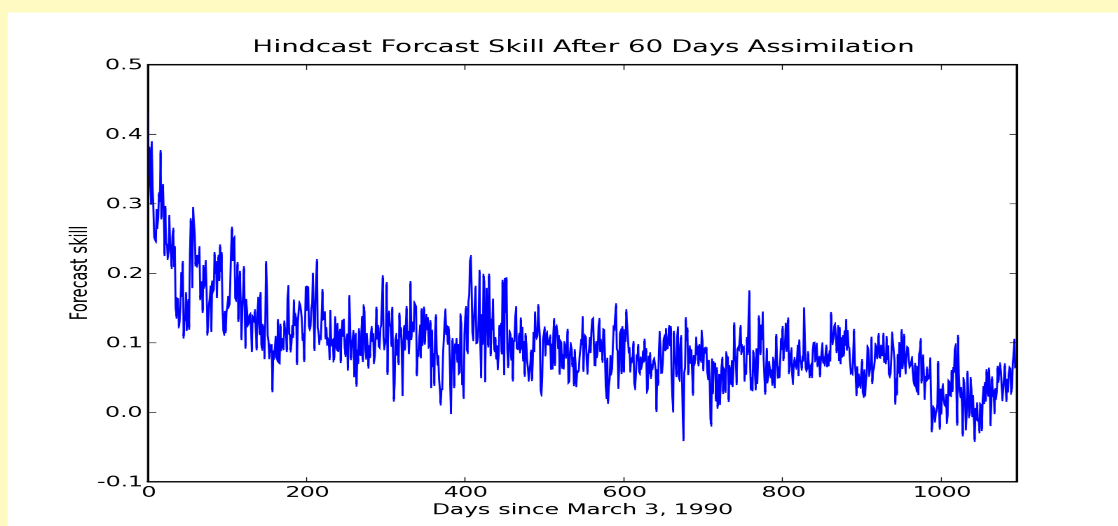
Improved Results





LEFT: new run SST bias (shows significant improvements over the control run. **RIGHT:** new run area-weighted mean absolute error. We observe a significant reduction in error, compared to the control run.

Hindcast Results



Hindcast results with the new inflation scheme show great improvement over the initial hindcast results.

Hindcast skill exhibits a fast decay over the first six months, followed by a slower decline. Results seem to demonstrate the presence of at least some level in forecast skill over the control for longer than a year.

Open Questions

- Better idea for fixing ensemble spread?
- Results with ARGO data?
- Assimilate averaged observations?
- AMOC improved with DA? Need to adjust localization?
- Decadal predictability?
- Higher resolution?

References

Anderson, J. (2009). Spatially and temorally varying adaptive covariance inflation for ensemble filters. *Tellus A*, 61:72-83.

Anderson et al. (2009). The Data Assimilation Research Testbed: A community facility. *Bull. Amer. Meteor. Soc.*, pages 1283-1296.

Griffies et al. (2009). Coordinated Ocean-ice Reference Experiments (COREs). *Ocean Modelling*, 26:1-46.

Large, W. G. and Yeager, S. (2009). The global climatology of an interannually varying air-sea flux data set. *Climate Dynamics*, 33:341-364.